DOCUMENT RESUME

ED 280 159 EA 019 268

AUTHOR Purser, Susan R.
TITLE The Relationship between Teacher Effectiveness and Teacher Evaluation and Selected Teacher Demographic Variables.
PUB DATE Feb 87
NOTE 22p.; Paper presented at the Annual Meeting of the American Association of School Administrators (New Orleans, LA, February 20-23, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Cultural Background; Grading; High Schools; Race; Rating Scales; Sex; Teacher Certification; *Teacher Effectiveness; *Teacher Evaluation; Teaching (Occupation); Teaching Experience
IDENTIFIERS Mississippi (Jackson County)

ABSTRACT

This paper inquires into the relationship between the criterion variable of teacher effectiveness and the independent variables of the score on the teacher evaluation procedure and the teacher demographic variables of race, sex, level of teacher certification, area of teacher certification, and years of teaching experience. The ultimate goal of the study was to provide data to assist school districts in improving the process of predicting and assessing teacher effectiveness. The study was conducted in an urban school district with a student population of 30,000 (70 percent minority) and a certified staff of 1,700 employees. High school teachers were classified by their race, sex, level of teacher certification within each of four subject areas (English, mathematics, science, and social studies). Effectiveness was rated by assigning teachers to standard, below standard, or above standard categories based on predicted final grades compared to actual final grades for the students of that teacher. Data were subjected to statistical analysis. Findings supported the hypothesis that there was no statistical relationship between teacher effectiveness and variables of race, sex, level of certification, area of certification, or years of experience. There also was no significant relationship between the score on the traditional teacher evaluation summative report and teacher effectiveness. A statistical data table and 15 references are included. (WTH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED280159

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Susan R. Purser

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

THE RELATIONSHIP BETWEEN TEACHER EFFECTIVENESS AND TEACHER EVALUATION AND SELECTED TEACHER DEMOGRAPHIC VARIABLES

Susan R. Purser, Ed.D.

Principal
Jackson Public Schools
P.O. Box 2338
Jackson, MS 39205

Paper presented at the Annual Meeting of the American Association of School Administrators (New Orleans, Louisiana, February 20-23, 1987).

EA 019 268

THE RELATIONSHIP BETWEEN TEACHER EFFECTIVENESS AND TEACHER EVALUATION AND SELECTED TEACHER DEMOGRAPHIC VARIABLES

Susan R. Purser, Ed. D.

Teacher effectiveness is a call word of the times. Teacher evaluation is a process that has been going on informally for generations. How do these two terms "fit" together? How should teacher effectiveness be evaluated? Have methods of today and yesterday done a bad job in determining who has an impact on students?

The term teacher effectiveness is being used quite frequently by the researchers as well as the public. Each use, however, seems to carry with it a different meaning. Even though the attempts at determining the distinguishing characteristics have not been completely successful, there is no reason to reach the conclusion that the questions of teacher effectiveness cannot be answered (Doyle, 1978).

Research in the area of teacher effectiveness has gone from observation and description to correlational studies to experimentation. Teaching practices were observed and recorded. Later those practices were correlated to student achievement. Now those same practices are being manipulated as determinations are being made about their impact on student achievement (Gage, 1984).

In 1978, school administrators indicated that from 5 to 15 percent of the teachers within their districts were not performing up to an acceptable level (Gudridge, 1980).

Edwin M. Bridges, in an editorial for Education Week, November 20, 1985, very pointedly summarized the impact:

> Incompetent teachers represent a relatively small proportion of the work force, but the number of students who are being taught by such teachers is substantial. Even if only 5 percent of the teachers in public elementary and secondary schools are incompetent, the number of students being taught by these teachers exceeds the combined public school enrollments of the 14 smallest states.

As a promise of hope for the field of education, one of the implications of research is that most ineffective teachers can be helped to improve when a proper evaluation tool is used (Rosenholtz, 1984).

The problem of this study is stated as the following question: What is the relationship between the criterion variable of teacher effectiveness and the independent variables of the score on the teacher evaluation procedure and the teacher demographic variables of race, sex, level of teacher certification, area of teacher certification, and years of teaching experience?

In order to clarify the problem that was studied, the following terms are briefly defined:

Teacher effectiveness. A teacher was placed in one of three categories based on a comparison of the actual and predicted final grades of the students in his/her classes. These categories were (a) standard level of effectiveness, (b) below standard level of effectiveness, and (c) above standard level of effectiveness.

Level of teacher certification. The level of certification was the classification that the State Department of Education assigned to a teacher.

Years of teaching experience. Years of teaching experience included the total number of years that the teacher had been employed as a teacher, regardless of location. The definition assumed that the teacher had been present for a minimum of 35 weeks to receive credit for a year of experience.

The general purpose of this study was to determine the accuracy of an existing performance evaluation system in determining the effectiveness of a teacher. In addition, the relationship among the teacher demographic variables of race, sex, level of teacher certification, area of teacher certification, and years of teaching experience, and the criterion variable of teacher effectiveness were examined. The ultimate goal of the study was to provide data which may assist school districts in improving the process of predicting and assessing teacher effectiveness.

The study was conducted in an urban school district with a student population of 30,000 which was approximately 70% minority. The certified staff included 1700 employees.

To answer the question raised by this study, the following hypotheses were tested:

1. There is a significant relationship between the predictor variables of the composite score on the teacher evaluation summative form, race, sex, level of teaching certification, area of teacher certification,

and years of teaching experience and the criterion variable of teacher effectiveness.

2. There is no significant relationship between the predictor variables of race, sex, level of teaching certification, area of teaching certification, and years of teaching experience and the criterion variable of teacher effectiveness.
3. There is no significant independent relationship between each of the predictor variables of the composite score on the teacher evaluation summative form, race, sex, level of teaching certification, area of teacher certification, and years of teaching experience, and the criterion variable of teacher effectiveness while holding constant the remaining variables.
4. There is a significant relationship between the scores in each of the performance areas on the teacher evaluation summative form and the criterion variable of teacher effectiveness.
5. There is no significant relationship between the predictor variables of race, sex, level of teacher certification, area of teacher certification, and years of teaching experience and the criterion variable of the composite score on the teacher evaluation summative form.

The problem of this study focused primarily on the relationship between a rating of teacher effectiveness and scores from a teacher evaluation procedure. The definition used for teacher effectiveness centered on student outcomes as determined by final grades. The evaluation procedure that was used by the school district relied on the judgement of the principal. The process included conferences which relate to classroom observations and was completed when the principal assigned ratings on the summative report. The procedure did not utilize student outcomes.

This study, therefore, examined the relationship that existed between what was happening in terms of student performance and what was observed by the principal to be happening. In other words, was the teacher evaluation procedure accurately identifying the teachers whose student outcomes were the most desired?

The questions being raised here are typical of those echoing through the halls of education and the halls of the nation's governing bodies. The talk is focusing on who is effective and how to determine such. The real issue, however, has to do with the fact that educators have not done a good job in policing the ranks. It is because of those who are not doing a good job that the image of education has become tarnished.

The impetus for studying effective teachers/teaching may have been to ward off the negative, but the results will aid education to move toward the positive.

Even though research about teaching has been taking place for many years, the Coleman Report of 1966, Equality of Educational Opportunity, caused a broad-based reaction and probe into action. The Coleman report and its offshoots have minimized the role of the teacher in accounting for educational outcomes. Many researchers have criticized the assumptions and the research designs and methods that have been utilized in most studies of teachers and teachings.

The first recorded study of teacher effectiveness was conducted by Kratz in 1896 (Medley, 1982). He asked a large group of elementary student to write down the characteristics that distinguished the best teacher they ever had from all of the other teachers. These descriptions were compiled, and from them a list of characteristics was derived. this list supposedly provided the basis for discriminating between effective and ineffective teachers. This process has been repeated many times, with and without modifications. The underlying assumptions are that anyone can judge teacher effectiveness and that these teachers were born, not made.

Early in this century, the teacher rating scale appeared. As early as 1930, Barr was advocating the use of student gains as the criteria of teacher effectiveness (Medley, 1979). He collected 209 different rating scales and attempted to validate such ratings by comparing their ratings by the supervisors with the average achievement gains of the pupils in sample classes. This study proved to be a failure. He found no correlation between the effectiveness ratings and the effectiveness measures.

Others have failed to validate the ratings of teacher effectiveness. No correlation between the ratings of teacher effectiveness have been found. Rosenshine did find, however, some support of a correlation between effectiveness measures and effectiveness ratings when trained observers were used. However, no evidence was found when supervisors were used to complete the rating instruments.

Looking at teacher characteristics that were hoped to be predictive of teacher effectiveness, the findings have been quite mixed (Ryans, 1958; Fattu, 1963).

1. There was only a slight relationship between the rated success of a teacher and his/her intelligence. The correlations were somewhat higher and less variable with high school teachers than with elementary.
2. Knowledge of subject matter appeared to be dependent upon the particular situation.
3. There was no substantial evidence that there is a relationship between the cultural background and his/her effectiveness.
4. The relationship between the teacher's socio-economic status and his/her effectiveness rating was low.
5. Married teachers do not appear to be less effective than their single colleagues.
6. The attitude toward teaching did show a small, positive relationship to teacher effectiveness.

7. Years of teaching experience seem to have a negative relationship. It appears that a positive correlation exists only until about the fifth year, then these is a leveling off for the next 15 to 20 years, and then a gradual decline in the effectiveness rating.
8. At the elementary level, sex does not appear to account for differences while at the secondary level, women as a group appear to be more effective than men.
9. No research was found which focused on the race of the teacher of the suject area taught.

With the 1960s came the process-product studies. The question was -- do students perform better because of the teacher or does the teacher behave in such a way because the students are performing? Therefore, these studies attempted to correlate measures of teacher performance with measures of teacher effectiveness.

Rosenshine and Furst (1976) provide one of the best syntheses of the process-product studies. They quickly caution, however, that the variables which they have extracted are correlational, not causal. These findings, more than anything, provide some hypotheses upon which to build further studies.

The tentative findings of Rosenshine and Furst yielded eleven variables; the best results were obtained from the first five:

1. Clarity of the teacher's presentation

2. Variability of the cognitive levels of discourse, the level of student tasks, and the classroom techniques and instructional materials.
3. Enthusiasm expressed by the teacher.
4. Task oriented and/or business like manner of the teacher.
5. Student opportunity to learn criterion material.
6. Use of student ideas and general instructions.
7. Criticism by the teacher.
8. Use of structuring comments, such as at the beginning or end of a lesson.
9. Types of questions asked by the teacher.
10. Probing responses by the teacher which encourage a student to elaborate on an answer.
11. Level of difficulty of instruction.

The methods used in the research projects which generated this list were usually of a rating scale or counting nature -- recording not evaluating.

Despite its shortcomings, process-product research can produce, and has produced, reliable information about the nature of effective teaching. When teachers are visited by observers trained to record the teacher behaviors accurately and objectively, appropriate analyses of the listings reveal differences between the behaviors of teachers who are more effective in helping pupils grow in basic skills and in some affective areas and the behaviors of those who are less effective (Medley, 1979).

Cook (1985) determined that the use of standardized achievement test scores in assessing teacher effectiveness is both possible and feasible. He suggested that such a factor utilizes an objective element in the assessment of teachers. There are others who argue that commercially prepared standardized tests should not be used to measure student achievement. They cite that the validity of the results using such instruments is questionable because such tests may not measure the teacher's instructional objectives. Norm referenced tests are designed to be independent from specific curricula and to draw from a student's total experience, not just from his formal instruction while in school. Perhaps the results from several comparable tests should be used to support a judgement of a teacher.

There are two potential problems in using student achievement in determining teacher effectiveness: the variability in student factors over which the teacher has no control and teachers teaching the test.

Medley (1979) highlights the focus for today's issue of teacher effectiveness with this statement:

> It seems logical, therefore, to assess teacher effectiveness on the basis of how much content a teacher's students master in how brief a period of time, with allowances made for relevant characteristics of the particular group of students being taught and other factors affecting pupils' learning that are beyond the teacher's control (p. 17).

In concurrence with this thought, Burton (1980) found that student background factors are important in their influence on

achievement. Among these factors were race, sex, and qualification for assistance with lunch purchase.

Medley (1979) contends that if teachers know the criteria on which decisions affecting their careers are based, they will meet the criteria if it is humanly possible to do so. This notion of controlling for certain background factors while delineating the expectations for teachers provides direction while exhibiting faith in the classroom professional.

Again the caution, the research that has taken place is descriptive and correlational and does not establish cause and effect. But while the tide appears to be returning to that of the late 19th Century in Britain when compensation was based on the performance by students (Travers, 1981), these studies do provide direction for future experimental studies.

As stated earlier, this study examined the relationship between an existing teacher evaluation system and a model for identifying effective teachers. A selected set of teacher demographic variables were examined for their predictive relationship with the variable of teacher effectiveness and their relationship to the evaluation score.

High school teachers were classified by their race, sex, and level of teacher certification within each of four (4) subject areas: English, mathematics, science, and social studies. After dividing the population of teachers by exact courses taught, 87 teachers were randomly selected from each combination of these same factors. An attempt was made to have as equal a number of teachers from each subject area as possible.

In order to determine the effectiveness category for these selected teachers, student achievement was used. The total district population of students enrolled in the identified courses was divided into eight (8) subgroups based on their race, sex, and socioeconomic status. Taking each course separately, an analysis of their final grades was conducted to determine the predictive value of the classification variables (race, sex, and socioeconomic status). A final grade was predicted for each subgroup within each course by using the generated predictive weights.

The averages of the actual and predicted final grades for the students taught by a teacher were calculated for each teacher in the sample. An interval was established around the average of

the predicted final grades. If the average of actual final grades for a teacher's students fell within the interval, the teacher was placed in the standard category. If the average were higher than the averages in the interval, the teacher was classified as above standard. The below standard group was composed of those teachers whose average of actual student final grades was less than the average within the interval. No quotas were involved in placing teachers in these categories.

After each teacher was assigned to an effectiveness group, all of the data were in place for the hypotheses to be tested. The statistical techniques of multiple linear regression and multiple discriminate function were used to analyze the data. The .05 level of statistical significance was used as the rejection level for all of the hypotheses.

The major findings can be summarized as follows:

1. There was no significant relationship between the categories of teacher effectiveness and the teacher variables of race, sex, level of teacher certification, area of teacher certification, years of teaching experience, and the score from the teacher evaluation procedure.

2. Collectively, the teacher variables of race, sex, level of teacher certification, area of teacher certification, and years of teaching experience had no significant relationship with the categories of teacher effectiveness.

3. Individually, the teacher variables of race, sex, level of teacher certification, area of teacher certification, years of experience, and the teacher evaluation score did not have a

significant relationship with the categories of teacher effectiveness.

4. There was no significant relationship between the scores of the performance areas (instruction, class management, interpersonal relations, and professional responsibilities) on the teacher evaluation report and the categories of teacher effectiveness.

5. The teacher variables of race, sex, level of teacher certification, area of teacher certification, and years of teaching experience did not have a significant relationship with the score from the teacher evaluation procedure.

This study indicated that there was no significance relationship between selected teacher demographic variables including the evaluation score and the categories of teacher effectiveness. It was anticipated that effectiveness was not a function of race, sex, and area of teacher certification. This expectation was verified.

The factors of the level of teacher certification and years of teaching experience were also not predictive of teacher effectiveness. These latter results concur with the findings of Ryans (1958): the extensiveness of general and/or professional education is not predictive of the teacher's effectiveness. Also, the years of teaching experience seem to have a negative relationship, after a point.

This study also produced similar findings as the study of Nesbit (1983): the teacher evaluation procedure did not evidence any discrimination on the basis of the teacher demographic

variables. Nesbit's study and this study were conducted in the same district but at different times and with different methodology. The lack of discrimination that was found in each case, however, is a pleasing outcome for the school district.

There was one point of surprise: the score on the teacher evaluation summative report did not have a significant relationship with teacher effectiveness. It was thought that there would be a significant relationship between the existing system, which was based on observation of teacher performance, and a measure of teacher effectiveness determined by student outcomes. Stow (1979) found a similar outcome using fourth grade mathematics teachers and the Iowa district's evaluation system. Her results showed that student gain scores did not correlate with the scores for teacher performance. Achievement scores were up, but the process of teacher evaluation was not validated.

A plausible explanation for the result of no significance in this study is found in the lack of variability in the evaluation scores. As indicated in the Table, the mean of the evaluation scores for the below standard group was higher than the other two groups. As a matter of fact, the average score for the above standard group was the lowest of the three. This situation highlights a flaw in the system. This study, however, did not provide a reason for such a discrepancy.

As the performance area scores were reviewed, another fact came to the surface: the interpersonal relations and professional responsibility areas had the higher averages for the total sample. Interestingly, these two areas were highest for

TABLE

MEAN AND STANDARD DEVIATION OF PERFORMANCE AREAS
OF TEACHER SAMPLE
BY EFFECTIVENESS CATEGORY

PERFORMANCE AREA	STANDARD	ABOVE STANDARD	BELOW STANDARD	TOTAL
INSTRUCTION				
MEAN	3.21	3.21	3.19	3.21
STANDARD DEVIATION	.22	.23	.19	.21
CLASS MANAGEMENT				
MEAN	3.26	3.12	3.12	3.20
STANDARD DEVIATION	.32	.27	.20	.29
INTERPERSONAL RELATIONS				
MEAN	3.37	3.40	3.44	3.39
STANDARD DEVIATION	.39	.35	.38	.37
PROFESSIONAL RESPONSIBILITIES				
MEAN	3.36	3.21	3.42	3.34
STANDARD DEVIATION	.35	.29	.31	.33
TOTAL INSTRUMENT				
MEAN	48.72	48.17	48.78	48.61
STANDARD DEVIATION	3.48	3.10	2.27	3.13

those teachers in the below standard category. The areas of greatest concern these days, instruction and class management, had higher averages for the standard group. Teachers in the above standard group scored equal to or lower than those in the other two categories on all areas. This finding certainly was not the one anticipated.

Another weakness of the model used in this study may have been a factor brought about by the low variability within the assignment of the students' final grades. The final grades were computed as the average of the nine week grades and the semester examinations. Eighty percent (80%) of the nine week grade was the average on the computer graded, district developed, criterion referenced tests. Teachers had control over the remaining twenty percent (20%).

Based on the results of this study, the following three (3) questions are raised because of the evaluation score/effectiveness category paradox:

1. Is there a flaw in the manner in which teachers assign student grades?

2. Are principals not skilled in distinguishing between the levels of teacher performance?

3. Is the evaluation instrument correlated with the teacher behaviors which promote optimal student achievement?

Perhaps improvement is needed in all three areas.

The model which was proposed in this study has attempted to use data which are available to school districts everywhere.

Students are held accountable for their grades. They are promoted, recognized for honors, and cited for reprimands based on the grades that they have earned. Those responsible for providing the opportunity for those same students should be held accountable on the same grounds. This model could be revised and be more fully developed for use as a portion of the evaluation of teachers.

Since this model provided for teachers to be placed in one of three (3) categories, there are some ramifications for such a technique to be used in a merit pay system. Again, there are changes that may need to be made, but this model may serve as a beginning point.

REFERENCES

Bridges, E. M. (1985). Coping with incompetent teachers. Education Week, V (12), 24.

Burton, W. P. (1980). The relationship between achievement and selected school, teacher, student and science program characteristics. (Abstract from Dissertation Abstracts International, 41, 2071).

Cook, D. (1985). An objective component for the evaluation of teaching. (Abstract from Dissertation Abstracts International, 46, 562).

Doyle, W. (1978). Interpreting teaching effectiveness research. Viewpoints in Teaching and Learning, 54 (4), 141-153.

Fattu, N. A. (1963). Research on teacher evaluation. The National Elementary Principal, 43, 24.

Gage, N. L. (1984). What do we know about teaching effectiveness? Phi Delta Kappan, 66 (2), 87-93.

Gudridge, B. M. (1980). Overview. In Teacher competency: Problems and solutions, AASA Critical Issues Report (pp. 5-8).

Medley, D. M. (1982). Teacher effectiveness. In C.W. Harris (Ed.), Encyclopedia of educational research (pp. 1894-1903). New York: MacMillan Co.

Medley, D. (1979). The effectiveness of teachers. In P. Peterson & H. J. Walberg (Eds.), Research on teaching (p. 16). Berkeley, California: McCutchan Publishing Co.

Nesbitt, L. W., Jr. (1983). A study of the relationships between teacher-related and evaluator-related variables and two measures of teaching performance. (Doctoral dissertation, University of Southern Mississippi, 1983). Dissertation Abstracts International, 45, 1004.

Rosenholtz, S. J. (1984). Conclusions. MYTHS: Political myths about reforming teaching, a paper prepared for the Education Commission of the States, (No. TQ84-4), p. 22.

Rosenshine, B. V. & Furst, N. (1976). Research on teacher performance criteria. In B. O. Smith (Ed.), Research in teacher education: A symposium, (p. 37). Englewood Cliffs, New Jersey: Educational Technology Publications.

Ryans, D. G. (1960). Predictions of teacher effectiveness. In C. W. Harris (Ed.), Encyclopedia of educational research (pp. 1486-1491). New York: MacMillan Co.

Stow, S.B. (1979). Using effectiveness research in teacher evaluation. Educational Leadership, 37, 55-58.

Travers, R. M. W. (1981). Criteria of good teaching. In J. Millman (Ed.), Handbook of teacher evaluation (pp. 14-22). Beverly Hills: Sage Publications.